

ユーザの既訪問スポットの位置づけに基づく未訪問スポットの説明手法

潘 健太[†] 北山 大輔[†]

[†] 工学院大学大学院工学研究科情報学専攻 〒163-8677 西新宿 1-24-2

E-mail: [†]tem18011@ns.kogakuin.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp

あらまし 近年、観光スポットを決める時に Web 上の観光情報を活用して計画を立てることが多くなっている。しかし、ユーザが多くエリアから訪問したいエリアを決めた上で、さらに自分のイメージに合う観光スポットを探すのは膨大な時間と労力を必要とする。また、ユーザが未訪問スポットに対して期待と不安を感じる場合がある。本研究では、ユーザの未知なスポットに対する理解を支援するためには、既に訪問したことがある観光スポットの特徴を未訪問スポットにあてはめて理解を支援する説明手法を提案する。観光スポット自身の特徴を重視するため、各観光スポットの特徴抽出に、ユーザが入力した観光スポットのすべてのレビュー、対象エリアの観光スポットのすべてのレビューを使用する。また、プロトタイプシステムを構築し、既訪問スポットと未訪問スポットとの説明情報の効果を検証する評価実験を行う。

キーワード 観光スポット, 理解支援, レビュー, コサイン類似度, TFIDF, 調和平均

1. はじめに

旅行先を決定する時、旅行者は観光スポット検索サイトや観光情報に関連する書籍を見て観光スポットを選び、旅行計画を立てる。しかし、ユーザにとって訪問したいエリアを決定した後、さらにエリア内に数多く存在する観光スポットから、自身のイメージから外れない観光スポットを見つけることは容易ではない。行きたい観光スポットが決まっていなかった場合はランキングやおすすめ情報を見て観光スポットを決めることが多くなると考えられる。この時、ユーザが選択した観光スポットに対するイメージが曖昧になるため不安を感じる場合がある。

近年、観光業とソーシャルネットワーキングサービスの発展スピードが加速しており、体験した観光スポットに対するレビューを観光スポット検索サイトに投稿しているユーザが増加している。さまざまな観光スポットを効果的に理解するためには、既存の情報をもとにして、未知な情報と既知な情報との対応関係を考えることが不可欠となる。この考え方は、以前に経験した事柄（ベースと呼ぶ）を、現在直面している事柄あるいは問題（ターゲットと呼ぶ）にあてはめる類推に相当する。たとえば、金沢の「にし茶屋街」という未知なスポットに対して既訪問の京都の「花見小路」と似ていると説明するとイメージの理解がしやすくなる可能性がある。

本研究では、ユーザの未知なスポットに対する理解を支援するため、既に訪問したことがある観光スポットの特徴を未訪問スポットにあてはめて理解を支援する説明手法を提案する。具体的には、ユーザが入力した既訪問スポットと未訪問エリアから、レビューを用いて既訪問スポット内の各スポットの独特な特徴と未訪問エリア内の各スポットの独特な特徴を抽出し、比較を行って説明情報を提示する。このプロトタイプシステムにより、ユーザが未訪問スポットに対する理解の支援を目指す。図 1 は提案手法の概念図である。

本論文の構成は下記のとおりである。2 節では関連研究につ

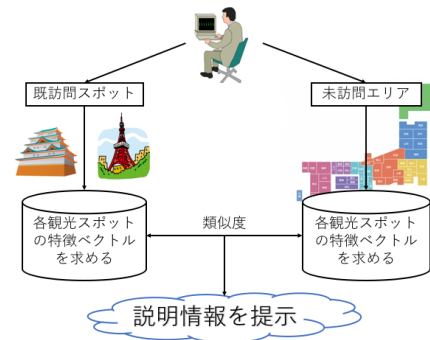


図 1 ユーザの既訪問スポットの位置づけに基づく未訪問スポット (エリア) の説明手法

いて述べる。3 節では提案手法の概要について述べる。4 節では構築したプロトタイプシステムの効果を検証する評価実験と考察について述べる。最後に 5 節ではまとめと今後の課題について述べる。

2. 関連研究

ユーザの体験履歴を利用した検索や推薦システムに関する研究が数多く発表されている。倉島ら [7] は、Flickr に投稿された写真のジオタグ情報を人々の旅行履歴として利用した旅行ルート推薦手法を提案した。この手法では、ユーザの現在地から行きやすい場所とユーザの興味に合致した場所に移しやすいと仮定し、行動モデルを生成している。ユーザのジオタグ付き写真集合は、時間情報でソートすると個人の旅行履歴とみなすことができると考え、ジオタグ情報を利用してユーザの行動モデルを生成している。北村らは [8]、一般的な物体認識を用いて、過去の個人旅行写真から旅行計画のユーザの嗜好を推定することに基づいて観光地を推薦する方法を提案した。物体認識システムを用いて、写真で撮った被写体情報のキーワードを取得し、

グラフ視覚化技術によってキーワードの共起を表現した。また、グラフの視覚化技術に基づいて旅行写真付きのグラフを視覚化するユーザインターフェイスを紹介した。Cheng らは [9], 自由に利用できるコミュニティ寄稿の写真を活用して、パーソナライズされた旅行のおすすめに焦点を当て、特定のユーザープロフィールまたは属性を考慮し、パーソナライズされた旅行の推奨を行うことを提案した。

類推は創造的思考に貢献すると指摘されてた [1]。既知の知識 (ベースと呼ぶ) から概念 (ターゲットと呼ぶ) を獲得するときには類推思考が働くと考えられる [2]。類推に関する研究の多くは、ベースとなる学習データとターゲットとなる問題が与えられ、物事の特徴を問題の特徴にマッピングして問題を解決するもの [3] である。Gick らは、不確定な問題の解を見つけるためのガイドとして、異種ドメイン間の類推の使用を調査するように設計した。学習データの与え方や機能について研究したもの [4] や、認知的な熟達度に応じて問題を解決するかどうかを明らかにしたもの [5] がある。これらを含む従来研究の多くにおいては、類推に用いるベースとターゲットを与えたうえで、一定の手順に従って問題解決を行っている。また、構造の類似性には 3 種類あり、特徴の共有数で決まる「対象レベルの類似性」、ベースに存在する関係とターゲットに存在する関係の共有度に基づく「関係レベルの類似性」、および題の解法あるいは目標レベルでの類似性である「プラグマティックな類似性」とがある。

従来のユーザの体験履歴を利用する手法では、履歴写真のジオタグ情報を分析し、ユーザの嗜好とする研究が多く行われている。また、類推技術に関して学習支援でよく使われている。本研究では、既訪問スポットと未訪問エリアのレビューを使って、ユーザ既訪問スポット集合と未訪問スポット集合のそれぞれ集合の各スポットの相対的な特徴を求め、関連付けることによって、スポットに対する理解を支援するため説明情報を提示することができる。また、本研究では、類推の質を明示的に扱うため、構造の類似性「関係レベルの類似性」に近いと考えられる。

3. 未訪問スポットの説明手法

我々は、ユーザの既訪問スポットの位置づけに基づく未訪問スポットの説明手法を提案する。具体的にはまず、ユーザが既訪問の複数個の観光スポットと訪問したい観光スポットエリア情報を入力する。既訪問スポットレビューベクトルを使って既訪問スポット毎の特徴ベクトルを求める。未訪問スポットも同様にエリア内の各スポットの特徴ベクトルを求める。次に、既訪問スポットレビューベクトルと未訪問スポットレビューベクトルの差分特徴に類似する特徴を持つ未既訪問観光スポット関連付けを行う。最後に、TFIDF を用いて未訪問スポットの理解支援のための説明手法を定義し、ユーザに提示する。

3.1 スポットのレビューから特徴ベクトル生成

既訪問スポットや未訪問スポットのレビューベクトルは、形態素解析器「mecab-ipadic-NEologd」^(注1)で分かち書き (原型)

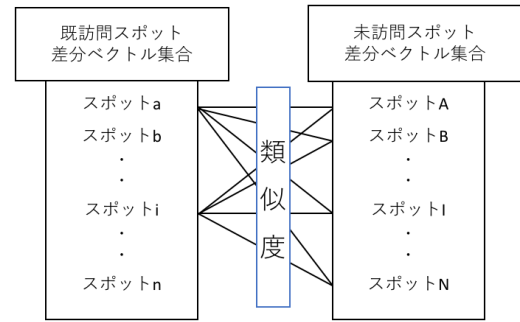


図2 類似度計算概念図

したレビューを利用して作成する。その後、Doc2Vec^(注2)の Distributed Bag-of-Words を利用して、各スポットの全レビューを使って 300 次元で作成したベクトルを使う。本稿に置いて、レビューデータは 2016 年 09 月末までじゃらん^(注3)から取得したものをを用いる。

3.2 観光スポットの役割に関する相対的な特徴

本研究では、観光スポットの特徴は相対的な特徴を利用する。相対的な特徴とは、特定の観光スポットが、ある観光スポット集合に含まれた他の観光スポットと比較した場合における独特な特徴である。例として、観光スポット集合内に鹿苑寺と清水寺が存在する場合を考える。このとき鹿苑寺の特徴は、金色、金箔、輝き等となり、清水寺の特徴は、舞台、胎内、一望等となる。どちらも京都に存在する寺院であるため、京都や寺院に関連する特徴は独特な特徴として現れることがない。次に、観光スポット集合内に東京都庁舎展望台と鹿苑寺が存在する場合を考える。このとき鹿苑寺の特徴は、金閣寺、お寺、金色、京都等となり、東京都庁舎の特徴は、展望、夜景、新宿等となる。観光スポットのカテゴリーが大きく異なる場合であれば、カテゴリーとしての特徴が現れる。また、スポット自身の特徴を表すことができる。本研究では、あるスポットが集合内の他のスポットと比較するとき、より各スポットの特徴を明らかにできる相対的な特徴に着目して研究を行う。

スポット差分ベクトルは式 1 として定義される。スポット差分ベクトルを求めるスポットを除いたスポット集合の各スポットのスポットベクトルの平均値を引いた値となる。 $spot_{set} = \{s_1, s_2, \dots, s_i, \dots, s_n\}$ は既訪問スポット集合や未訪問スポット集合となっている。また、 s_i は集合内のある観光スポットを示している。

$$v_i = s_i - average(spot_{set} - s_i) \quad (1)$$

3.3 説明スポットの決定

既訪問スポットの各特徴差分ベクトル v_i と未訪問スポットの各特徴差分ベクトル v_j から、既訪問スポットと未訪問スポット間の相対的な特徴の類似度 (図 2) を求める。類似度計算に

(注2) : <https://radimrehurek.com/gensim/models/doc2vec.html>

(注3) : <https://www.jalan.net/kankou/>

(注1) : <https://github.com/neologd/mecab-ipadic-neologd/>

表 1 形態素解析の例

レビュー文書	園内も広く、気分転換に散歩したりするのにちょうどよい。きれいに清掃などもされていて、気分がよいです。
形態素解析	園内 広い 気分転換 散歩 ちょうど よい きれい 清掃 されい 気分 よい

は、コサイン尺度 (式 2) を用いる。

$$\cos(v_i, v_j) = \frac{v_{i1}v_{j1} + v_{i2}v_{j2} + \dots + v_{in}v_{jn}}{\sqrt{v_{i1}^2 + \dots + v_{in}^2} \times \sqrt{v_{j1}^2 + \dots + v_{jn}^2}} \quad (2)$$

既訪問スポットの各特徴ベクトルと未訪問エリア内の各特徴ベクトルの類似度が 0.125 以上かつ、類似度が最も高い既訪問スポットと未訪問スポットの関連付けを行う。また、既訪問スポットと未訪問スポットとの関連をつけるとき以下の 2 つの方法がある。

(1) 既訪問スポットをベースにして未訪問スポットと関連付ける方法

(2) 未訪問スポットをベースにして既訪問スポットと関連付ける方法

方法 1 では、既訪問スポットをベースにすると未訪問エリア内のスポットが複数の特徴を保持する場合は同じ未訪問スポットと関連づける場合がある。方法 2 では、未訪問スポットをベースにすることによって、既訪問スポット集合から各スポットの特徴を取り出し、未訪問スポットと関連付けることができるため、方法 2 を利用する。また、本研究では、未訪問スポットを説明するため未訪問スポットをベースにすることが妥当と考えられる。

3.4 説明するための役割語の抽出

観光スポットのレビューはすべて形態素解析器「mecab-ipadic-NEologd」を使用することで、単語抽出処理を行う。しかし、これらを用いて得られた単語は、日本語として成立しない語が含まれており、これらノイズの削除が必要となる。具体的には、助詞、助動詞、連体詞、記号を削除する (表 1)。

節??で関連付けした既訪問スポットと未訪問スポットの情報は単語形式でユーザーに提示するため、ある観光スポットのレビュー集合を文書 i とし、 i に対する単語 j が出現するスポット集合の出現回数を $TF_{i,j}$ 、単語 j がスポット集合の文書数を DF_j 、スポット集合内の全スポット数を $|D|$ とした時、そのスポットにおける単語の特徴量は、式 3 で定義される。

$$word_{i,j} = TF_{i,j} \times IDF_j \quad (3)$$

$$IDF_j = \log\left(\frac{|D|}{DF_j}\right) \quad (4)$$

本手法では、既訪問スポットに関して、ユーザーが複数個のスポットを入力する。それぞれのスポットの全レビューをまとめて 1 つの文書と見なし、それ以外のスポットの全レビューも文書とみなすことで、式 3, 4 によって TFIDF 値を算出し、既訪問スポット毎の特徴語とする。

未訪問エリアに関して、ユーザーがエリアを指定して入力する。エリア内のそれぞれのスポットの全レビューをまとめて 1 つの

文書と見なし、それ以外のスポットの全レビューも文書とみなすことで、式 3, 4 によって TFIDF 値を算出し、未訪問エリアのスポット毎の特徴語とする。

関連付けした既訪問スポットと未訪問スポットの説明情報は、TFIDF で求めた各スポットの特徴語による調和平均を用いて決定する。調和平均とは、逆数の算術平均の逆数である。既訪問スポットのレビュー文書と、未訪問スポットのレビュー文書に、共通して出現する単語を抽出する。抽出した単語のスコアは式 5 によって定義する。 $word_{visited}$ と $word_{unvisited}$ は同じ単語がそれぞれ既訪問スポットの TFIDF 値と未訪問スポットの TFIDF 値を示している。単語スコアの値が大ききと既訪問スポットと未訪問スポットのそれぞれの TFIDF 値が大きき、つまり単語がそれぞれの文書に置いて重要度が高いことを示している。よって、単語スコアの上位 10 個の単語を説明情報としてユーザーに提示する。

$$score = \frac{1}{\frac{1}{2}\left(\frac{1}{word_{visited}} + \frac{1}{word_{unvisited}}\right)} \quad (5)$$

3.5 未観光スポットの説明情報の例

4. 評価実験

4.1 関連を表す特徴語の実験

節??では、調和平均を利用して単語スコアを算出し、上位 10 個の単語を説明情報としてユーザーに提示するの提案手法について、妥当性を調査する実験を行った。

4.1.1 実験内容

調和平均を利用して類推情報を提示する場合と、相加平均を利用する場合の比較を行った。まず、訪問スポットのレビュー文書と、未訪問スポットのレビュー文書に、共通して出現する単語を抽出する。次は、

閾値を決めて各スポットの TFIDF 値の平均以上の単語を抽出する。具体的に、既訪問スポット集合の各スポットの TFIDF 値の平均を計算し、スポット毎の単語の TFIDF 値が各スポットの平均の重み以上であるかどうか判断する。また、未訪問スポットも同様な算出方法を使つての集合の各スポットの TFIDF 値の平均を計算し、スポット毎の単語の TFIDF 値が各スポットの平均の重み以上であるかどうか判断する。既訪問スポット単語スコアと未訪問単語スコアの差の絶対値が 0 に近いと、それぞれのスポットにおいての単語の意味合いが近いと示している。よって、差の絶対値が 0 に近い 10 個の単語を類推情報としてユーザーに提示する。

4.1.2 実験結果と考察

4.2 対応付けの比較実験

節??では、既訪問スポット集合と未訪問スポット集合それぞれの各スポットベクトルと他のスポットベクトルを使って類似度を計算することで観光スポットの相対的な特徴を求める手法について、妥当性を調査する予備実験を行なった。

4.2.1 実験内容

既訪問スポット集合と未訪問スポット集合それぞれの各スポットベクトルと他のスポットベクトルを使って類似度を計算する場合と、節??の既訪問スポット集合と未訪問スポット集合

表 2 既訪問スポット集合と未訪問スポット集合

既訪問スポット	未訪問スポット
鹿苑寺（金閣寺）	皇居東御苑
八坂神社	新宿御苑
清水寺	東京都庁展望室
龍安寺	浅草寺
伏見稲荷大社	明治神宮

表 3 特徴ベクトルや特徴差分ベクトルを利用する時の実験結果

特徴ベクトルを利用		特徴差分ベクトルを利用	
既訪問	類似未訪問 スポット	既訪問	類似未訪問 スポット
鹿苑寺	明治神宮	鹿苑寺	新宿御苑
八坂神社	明治神宮	八坂神社	明治神宮
清水寺	浅草寺	清水寺	浅草寺
龍安寺	明治神宮	龍安寺	皇居東御苑
伏見稲荷大社	明治神宮	伏見稲荷大社	明治神宮

それぞれの各スポットベクトルと他のスポットベクトルの差分を使って類似度を計算する場合のスポットの比較を行った。表 2 は、実験で使う既訪問スポット集合と未訪問スポット集合である。既訪問スポット集合は京都の寺院・神社のカテゴリーから 5 つの観光スポットを選んだ。未訪問スポット集合は東京都に存在する観光スポットから複数のカテゴリーに渡って 5 つの観光スポットを選んだ。

4.2.2 実験結果と考察

表 3 は実験結果となっている。特徴ベクトルを利用する場合、既訪問スポットのカテゴリーは寺院・神社であるため、類似する未訪問スポットもカテゴリーによって影響される。未訪問スポットの 5 つのスポットからカテゴリーが一致するスポットが類似スポットになる結果になっている。

一方、節??の提案手法を利用する場合、既訪問スポット集合内の鹿苑寺や龍安寺と他のスポットと比較するとき庭園という相対的な特徴を見つけることができる。同様に、未訪問スポット集合内の新宿御苑や皇居東御苑と他のスポットと比較するとき庭園という相対的な特徴を見つけることができる。結果、鹿苑寺と新宿御苑、龍安寺と皇居東御苑を関連づけることができる。また、鹿苑寺の庭園特徴に関して、同じく庭園特徴の新宿御苑と関連づける理由として鹿苑寺は庭園中に金閣が建っているに対して、新宿御苑は庭園中に旧御涼亭が建っていることから、庭園以外の要素も考量して類似性を求めていることがわかった。よって、既訪問スポット集合と未訪問スポット集合それぞれの各スポットベクトルと他のスポット差分ベクトルの方はより観光スポットの相対的な特徴を求めることができる。

5. まとめと今後の課題

謝 辞

文 献

- [1] K. J. Holyoak and P. Thagard, “Mental Leaps: Analogy in Creative Thought, MIT Press”, Journal of Japanese Society

- for Artificial Intelligence, Vol.11, No.3, pp.489, 1996
- [2] D. Gentner, “Structure-Mapping: A Theoretical Framework for Analogy”, Cognitive Science, Vol.7, pp.155–170, 1983
- [3] M. L. Gick and K. J. Holyoak, “Analogical Problem Solving”, Cognitive Psychology, Vol.12, pp.306–355, 1980
- [4] M. L. Gick and K. J. Holyoak, “Scheme Induction and Similarity in Analogical Transfer”, Cognitive Psychology, Vol.15, pp.1–38, 1983
- [5] Z. Chen and M. W. Daehler, “Positive and Negative Transfer in Analogical Problem-solving by 6-years-old Children”, Cognitive Development, Vol.4, No.4, pp.327–344, 1989
- [6] K. J. Holyoak and P. Thagard, “Analogical Mapping by Constraint Satisfaction”, Cognitive Science, Vol.13, pp.295–355, 1989
- [7] T. Kurashima, T. Iwata, G. Irie and K. Fujimura., “Travel route recommendation using geotags in photo sharing sites”, CIKM ’10 Proceedings of the 19th ACM international conference on Information and knowledge management, pp.579–588, 2010
- [8] R. Kitamura and T. Itoh, “Tourist Spot Recommendation Applying Generic Object Recognition with Travel Photos”, ITE Tech. Rep., Vol.42, No.12, AIT2018-94, pp.185–188, 2018
- [9] A. J. Cheng, Y. Y. Chen, Y. T. Huang and Winston H. Hsu, “Personalized Travel Recommendation by Mining People Attributes from Community-Contributed Photos”, MM ’11 Proceedings of the 19th ACM international conference on Multimedia, pp.83–92, 2011
- [10] T. Kudo, K. Yamamoto and Y. Matsumoto, “Applying Conditional Random Fields to Japanese Morphological Analysis”, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230–237, 2004